Clasificadores de Bayes

Daniel Otero Fadul

Departamento de Ciencias
Escuela de Ingeniería y Ciencias

Inteligencia Artificial

IA significa "Inteligencia Artificial". Se refiere a la capacidad de las máquinas o programas de computadora para realizar tareas que, típicamente, requieren inteligencia humana. Estas tareas incluyen el aprendizaje, la percepción, el razonamiento, la resolución de problemas y la comprensión del lenguaje natural.

Inteligencia Artificial

Otra definición de IA es en la que se considera esta área como el estudio de agentes que reciben percepciones del entorno y realizan acciones. Un agente es cualquier sistema que perciba su entorno mediante sensores y actúe sobre éste mediante actuadores.

Inteligencia Artificial

En general, IA es una vasta área en la que podemos encontrar una gran diversidad de técnicas con las que se le puede otorgar a un agente un conjunto de habilidades que le permiten llevar a cabo una tarea que normalmente es realizada por un humano.

Clasificadores de Bayes “ingenuos”

Los clasificadores de Bayes ingenuos son una familia de clasificadores que se basan en el teorema de Bayes para asignarles clases $\{C_k\}_{k=1}^{K}$ a un conjunto de vectores de características x:

$$(x_1, x_2, \dots, x_n) \to C_k.$$

Estos clasificadores se les consideran “ingenuos” porque asumen que las características del vector x son condicionalmente independientes dada una clase C_k.

Clasificadores de Bayes “ingenuos”

Hay tres tipos de clasificadores:

- **Bernoulli:** Cada una de las componentes del vector x es una variable binaria.
- **Multinomial:** Cada componente de x puede tomar un determinado conjunto de valores.
- **Gaussiano:** Se asume que cada componente de x es una variable continua cuya distribución condicional dada una clase C_k es una distribución normal.

Probabilidad Condicional

Sean A y B dos eventos que pertenecen a un espacio muestral Ω. La probabilidad de que ocurra el vento A dado que se sabe que el evento B ocurrió se define como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Si A y B son independientes, se tiene que $P(A|B) = P(A)$, lo que implica que $P(A \cap B) = P(A)P(B)$.

Probabilidad Total

Sea $\{A_i\}_{i=1}^n$ una colección de eventos mutuamente excluyentes que forman una partición del espacio muestral: $A_i \cap A_j = \emptyset$, $\forall i \neq j$ y $\cup_{i=1}^n A_i = \Omega$. Entonces, para cualquier evento $B \subset \Omega$, tenemos que

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n). \end{aligned}$$

Nótese que $\forall i, \; P(A_i) > 0$.

Teorema de Bayes

Sea $\{A_i\}_{i=1}^n$ una colección de eventos mutuamente excluyentes que forman una partición del espacio muestral: $A_i \cap A_j = \emptyset$, $\forall i \neq j$, $\cup_{i=1}^n A_i = \Omega$ y $P(A_i) > 0$ $\forall i$. Entonces, para cualquier evento $B \subset \Omega$ tal que $P(B) > 0$, tenemos que

$$
\begin{aligned}
P(A_i|B) &= \frac{P(B|A_i)P(A_i)}{P(B)} \\
&= \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}.
\end{aligned}
$$

Teorema de Bayes

Cuando la partición de Ω es de solo dos elementos ($A \cup \bar{A} = \Omega$), tenemos el siguiente caso especial:

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \end{aligned}$$

Teorema de Bayes

Papá Noel empaca los juguetes que reparte en Navidad en cajas que contienen veinte juguetes. Suponga que el 60% de todas las cajas que reparte Papá Noel no contienen juguetes defectuosos, 30% contienen sólo un juguete defectuoso, y 10% contienen dos juguetes defectuosos. Una caja es seleccionada al azar y se escogen dos juguetes aleatoriamente. Los juguetes seleccionados no son defectuosos.

- ¿Cuál es la probabilidad de que no hayan juguetes defectuosos en la caja seleccionada?
- ¿Cuál es la probabilidad de que sólo exista un juguete defectuoso en la caja seleccionada?
- ¿Cuál es la probabilidad de que haya dos juguetes defectuosos en la caja seleccionada?

Teorema de Bayes

Sean C_0, C_1 y C_2 los eventos de caja con cero juguetes defectuosos, un juguete defectuoso y dos juguetes defectuosos, respectivamente. Sea J_0 el evento de sacar dos juguetes sin defectos de una caja. Entonces, para responder la primera pregunta, necesitamos calcular $P(C_0|J_0)$:

$$\begin{aligned} P(C_0|J_0) &= \frac{P(J_0|C_0)P(C_0)}{P(J_0|C_0)P(C_0)+P(J_0|C_1)P(C_1)+P(J_0|C_2)P(C_2)} \\ &= \frac{(1)\left(\frac{6}{10}\right)}{(1)\left(\frac{6}{10}\right)+\left(\frac{19}{20}\frac{18}{19}\right)\left(\frac{3}{10}\right)+\left(\frac{18}{20}\frac{17}{19}\right)\left(\frac{1}{10}\right)} \\ &= 0.631229... \end{aligned}$$

Clasificador de Bayes Ingenuo

Sea $x = (x_1, x_2, \ldots, x_n)$ un vector de características y sea $\{C_k\}_{k=1}^{K}$ un conjunto de clases. Cuando utilizamos el clasificador de Bayes como nuestro modelo es necesario calcular la siguiente probabilidad:

$$\begin{aligned} P(C_k|x) &= \frac{P(x|C_k)P(C_k)}{P(x)} \\ &= \frac{P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k)}{P(x)}. \end{aligned}$$

Clasificador de Bayes Ingenuo

Por la regla de la multiplicación tenemos que el numerador es igual a

$$P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k) = P(C_k)P(x_1|C_k)\cdots P(x_n|x_1 \cap x_2 \cap \cdots \cap C_k).$$

Si las características son mutuamente y condicionalmente independientes dada una clase C_k, se tiene que

$$P\left(x_i|x_1 \cap x_2 \cap \cdots \cap x_{i-1} \cap x_{i+1} \cap \cdots \cap x_n \cap C_k\right) = P(x_i|C_k).$$

Por lo tanto, la primera expresión se puede reescribir de la siguiente forma:

$$P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k) = P(C_k)P(x_1|C_k)\cdots P(x_n|C_k).$$

Esta condición normalmente no ocurre en la vida real, razón por la cual el clasificador de Bayes se le considera "ingenuo". Sin embargo, este método se puede comportar bastante bien en diversas situaciones.

Clasificador de Bayes Ingenuo

Para determinar a qué clase pertenece algún vector x se realiza el siguiente cálculo:

$$\hat{y} = \underset{k=1,2,\ldots,K}{\operatorname{argmax}} P(C_k)P(x_1|C_k)\cdots P(x_n|C_k).$$